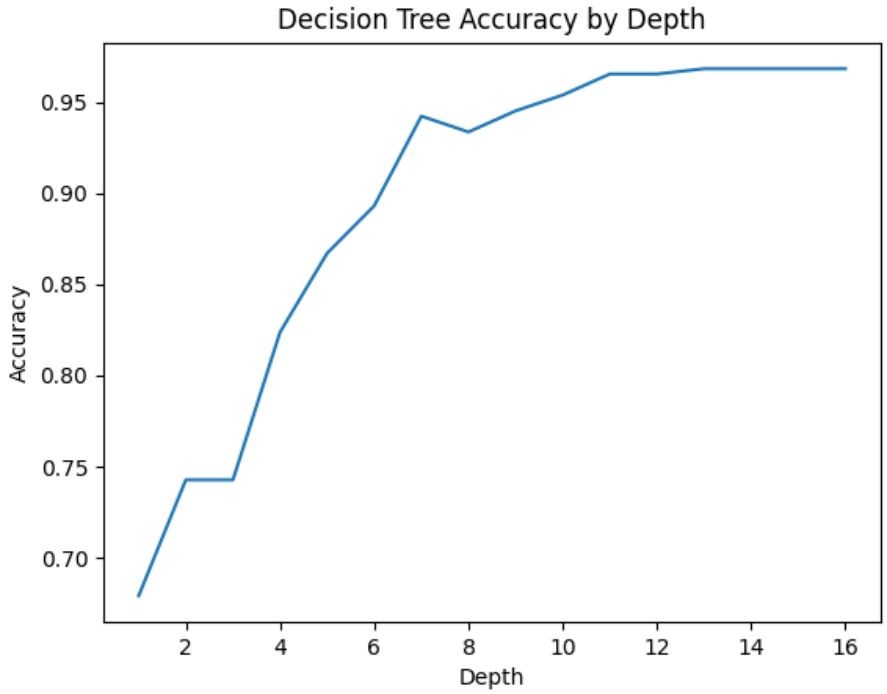
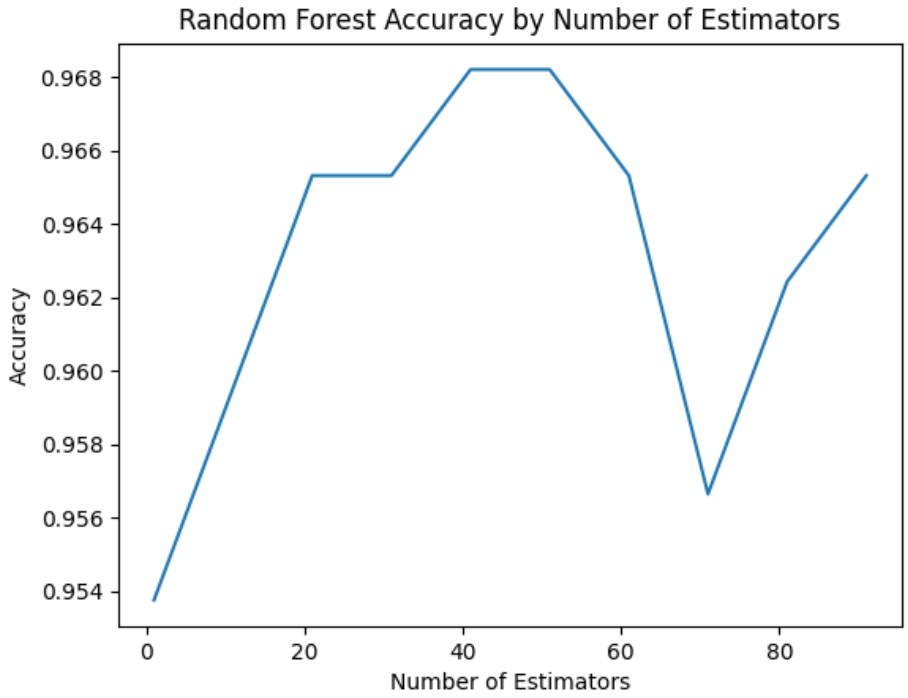
HW 5 Report

Decision Tree Accuracy vs Depth



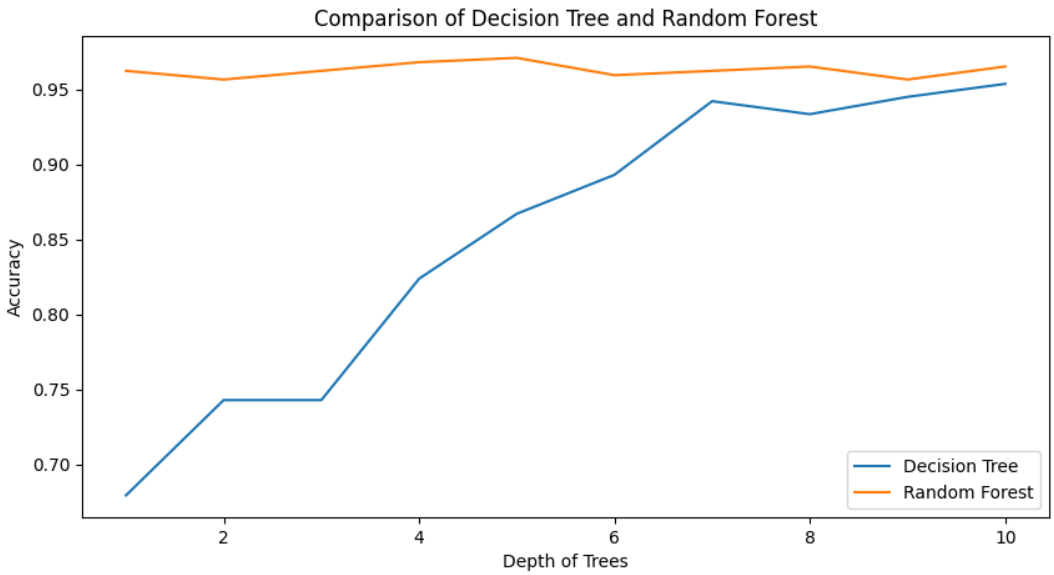
As the depth of the decision tree increases, the accuracy generally increases up to a point, as the model identifies the strongest patterns first. As weaker patterns begin to be used, the model plateaus off until there are no more splits that result in information gain. In this example, the model achieved a maximum accuracy of 96.8% at a depth of 13; however, at a depth of 7, the model achieved 94.2% accuracy, within 3% of the max at half the depth.

Random Forest Accuracy vs Estimators



As the number of estimators in the random forest increases, the accuracy will generally increase and begin to level off. Averaging the outputs of an increasing number of models reduces the variance and improves the accuracy. The base regression tree already performs well at an accuracy of 95.4%. This quickly improves to 96.5% after just 20 estimators and the maximum accuracy is reached After 40 estimators at 96.8%. After 40 estimators, the model sees no benefit in increasing the estimators, as it has already stabled, and more training could possibly lead to overfitting.

Decision Tree vs Random Forest



The random forest outperforms the decision tree at all depths for this scenario. Understandably, the decision tree starts out with poor accuracy at low depths because it can only predict off a few patterns. This differs from the random forest which has high accuracy above 95% even at shallow depths. The random forest creates trees that make predictions using subsets of the features, enabling them to learn trends that the decision tree cannot see early on. The random forest's performance remains consistent as the depth increases, and the decision tree eventually approaches the same accuracy. This demonstrates the strength of the random forest which can achieve high accuracy at low depths. It is also clear that there is little to no benefit in using deep trees for random forests, as there is no clear performance gain at a major increase in computation.